*The Biochemical Journal, Vol.* 128, *No.* 4

DR. C. B. ANFINSEN

Biochem. J. (1972) 128, 737–749
Printed in Great Britain

# The Formation and Stabilization of Protein Structure

## THE SIXTH JUBILEE LECTURE

By C. B. ANFINSEN*

*Delivered at a Meeting of the Biochemical Society on 28 March 1972 at University College London, Gordon Street, London WC1 0AH, U.K.*

Studies of protein molecules generally fall into one of three broad categories: polypeptide chain assembly from free amino acids, the folding of the chain into a unique three-dimensional object and the relationships between detailed geometry in solution and biological function. The last two, together with information on the molecular mechanism of enzymic catalysis, constitute the *raison d'être* of protein chemistry. The ultimate aim of the enzymologist and the protein chemist is to be able to synthesize an amino acid sequence that, when allowed to fold, will assume a stable predesigned three-dimensional arrangement of atoms capable of carrying out the desired catalytic act. It is a great honour to have been asked by the Biochemical Society to discuss some of the work that my colleagues and I have done over the past years that might contribute to the ultimate achievement of this goal.

The first category above, involving the assembly of amino acids into the sequence specified by the genetic material of the cell, is by now almost solved. It was the direct or indirect subject matter of the earlier Lectures in this series given by Paul Zamecnik (1962) and by Ghobind Khorana (1968), and remains of major interest to biochemists and molecular biologists who continue to fill in the details of nucleic acid interactions and metabolic control. We now know a great deal about how ribosome–mRNA complexes lay down the polypeptide chain from the *N*-terminus to the *C*-terminus, at a rate of about 1 residue/s per site in higher organisms and about one or two orders of magnitude more rapidly in micro-organisms.

The other two aspects of the study of proteins, namely folding and function, are very closely related. As our sophistication has increased it has become apparent that not much superficial dross remains in the fabric of proteins. Evolution, through natural selection based on function, has eliminated most of what is not needed (although it is difficult to be categorical about this; experiments carried out *in vitro* suggest that certain portions of structure may be eliminated without loss of function, but these may fulfil subtle and essential functions *in vivo*). Although most of the amino acids in the chain participate in directing folding and stabilization, certain proteins may be broken into inactive and structurally disorganized fragments that are capable of interacting with one another to regenerate active recombinants. In two fortunate instances, bovine pancreatic ribonuclease (Wyckoff *et al.*, 1970; Kartha *et al.*, 1967) and staphylococcal nuclease (Fig. 1), the recombined fragments form three-dimensional structures that have been shown crystallographically (Taniuchi *et al.*, 1972) to be essentially identical with the parent proteins (Arnone *et al.*, 1969, 1971). Since it is now possible to prepare fairly large peptide fragments by organic synthesis, the study of crystallizable analogues of these two proteins, containing one natural and one synthetic fragment, offers a unique opportunity to examine structure–function relationships at the atomic level. Such studies are now under way, and I would like to describe the groundwork that is being carried out in our laboratory and elsewhere in this direction.

### *Historical background*

Most proteins consist of one or more polypeptide chains that are devoid of disulphide-bond cross-links. The phenomenon of cross-linking, whether through disulphide bonds or through other more specialized types that exist, for example, in the fibrous proteins collagen and elastin, is found mainly in the extracellular proteins or structural proteins, which require greater stability to fluctuating environmental conditions. The reversibility of denaturation (which we will equate here in a rough way with partial or complete unfolding) has been under investigation and discussion for many years, and Anson (1945) and A. E. Mirsky, Lumry & Eyring (1954) and others had discussed the likelihood of such reversibility as long as 25 years ago. Because of the inability of the available hydrodynamic and spectral techniques to detect more than fairly gross changes in macromolecular structure, the spontaneity and precision of protein chain folding was not easily documented until studies were carried out on a variety of disulphide-bonded structures. In these cases one could detect, by the chemical identification of the half-cystine residues making up each bond, deviation

---

* Address: Laboratory of Chemical Biology, National Institute of Arthritis and Metabolic Diseases, National Institutes of Health, Bethesda, Md. 20014, U.S.A.

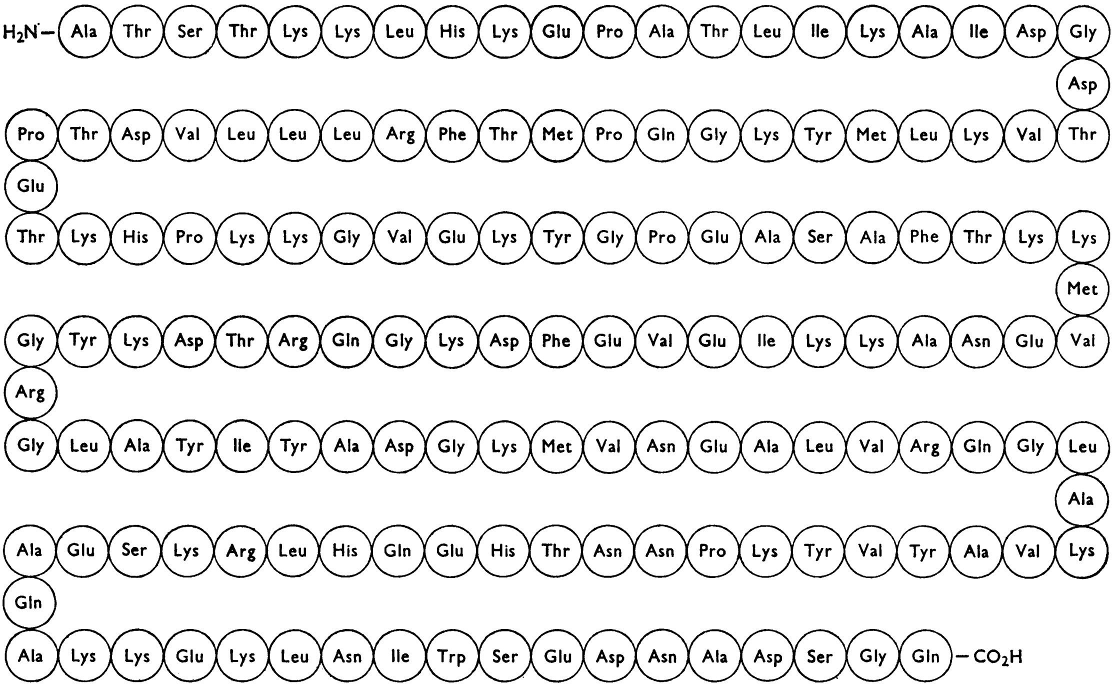

H₂N–
Ala Thr Ser Thr Lys Lys Leu His Lys Glu Pro Ala Thr Leu Ile Lys Ala Ile Asp Gly
Asp
Thr Val Lys Leu Met Tyr Lys Gly Gln Pro Met Thr Phe Arg Leu Leu Leu Val Asp Thr Pro
Glu
Thr Lys His Pro Lys Lys Gly Val Glu Lys Tyr Gly Pro Glu Ala Ser Ala Phe Thr Lys Lys
Met
Val Glu Asn Ala Lys Lys Ile Glu Val Glu Phe Asp Lys Gly Gln Arg Thr Asp Lys Tyr Gly
Arg
Gly Leu Ala Tyr Ile Tyr Ala Asp Gly Lys Met Val Asn Glu Ala Leu Val Arg Gln Gly Leu
Ala
Lys Val Ala Tyr Val Tyr Lys Pro Asn Asn Thr His Glu Gln His Leu Arg Lys Ser Glu Ala
Gln
Ala Lys Lys Glu Lys Leu Asn Ile Trp Ser Glu Asp Asn Ala Asp Ser Gly Gln
–CO₂H

Fig. 1. *Amino acid sequence of staphylococcal nuclease (Foggi strain) (Taniuchi & Anfinsen,* 1971; *Bohnert & Taniuchi,* 1972)

from correct pairing of thiol groups. In the case of ribonuclease, for example, the reduced chain containing eight half-cystine residues could be shown to form, on oxidation in air, only one of the 105 possible sets of four disulphide bridges (Sela *et al.*, 1957; White, 1961; Anfinsen & Haber, 1961). These and similar studies on a variety of other disulphide-bond-containing proteins indicated that the three-dimensional structure of the native protein in a given environment (solvent, pH, ionic strength, presence of other non-protein components, temperature etc.) is determined by the amino acid sequence and is the one in which the Gibbs free energy of the system is lowest (Epstein *et al.*, 1963). This environment, in our considerations here, is the physiological fluid of the living cell.

The fact that many of the earlier studies of protein renaturation and behaviour in solution were carried out with the more easily available disulphide-bonded proteins may possibly have introduced some minor misconceptions into the literature of protein chemistry. The so-called 'motility', or flexibility in solution, is quite marked for those that do not contain stabilizing cross-linkages, or prosthetic groups or other ligands that help rigidify the structure. Myoglobin, with its haem group, and ribonuclease, with its four disulphide bridges, contain a considerable number of backbone amide hydrogen atoms that are essentially inaccessible to bulk solvent as measured by the techniques of tritium or deuterium exchange. On the other hand, apomyoglobin and the disulphide-bond-free protein staphylococcal nuclease undergo almost complete exchange of all potentially exchangeable hydrogen atoms in a relatively short time (Schechter *et al.*, 1968, 1969).

The understanding of the kinetic and thermodynamic features of the folding of polypeptide chains into the specific geometry of protein molecules was also somewhat coloured by the complexities introduced by these cross-linkages. In our own early studies of the reoxidation of the reduced chain of ribonuclease, for example, the sluggishness of folding *in vitro* to yield the globular protein containing properly paired half-cystine residues suggested to us that incorrect pairing might be occurring, requiring subsequent reshuffling (Anfinsen *et al.*, 1961). Since it had been established that protein chains were synthesized, in the tissues of higher organisms at least, at a rate of at least 1 residue/s per synthesizing site (Dintzis, 1961; Canfield & Anfinsen, 1963), it was clearly out of the question to accept a process of folding with an overall duration of more than a very few minutes. The discovery of an enzyme in the 'microsomes' of a number of tissues (the endoplasmic reticulum, with or without attached ribosomal particles) that catalysed disulphide interchange in 'scrambled' protein molecules to yield the native structures helped to clarify the differences in rates between folding of disulphide-bonded proteins *in vivo* and *in vitro* (Goldberger *et al.*, 1963; Venetianer & Straub, 1963, 1964; Givol *et al.*, 1965). This enzyme, having a molecule weight of 42000 and containing three half-cystine residues, two in disulphide linkage and the third obligatorily in the thiol form for enzymic function (DeLorenzo *et al.*, 1966; Fuchs *et al.*, 1967), is associated with the membrane portion of the endoplasmic reticulum. Indeed, Rabin and his associates (Rabin *et al.*, 1970) have developed an ingenious assay for the occupation of the surface of the membrane with ribosomal particles on the basis of the disulphide-interchange enzyme. Its activity can be detected only when the enzyme has been 'revealed' by removal of the ribosome by, for example, certain steroids, carcinogens and metabolic poisons. The enzyme is most highly concentrated in the membranes of those metazoan tissues that are particularly active in the synthesis and secretion of disulphide-bond-containing proteins. It is tempting to speculate that this enzyme occupies a position in the structure of the cell membrane that permits its ready and perhaps essential involvement in the secretion of correctly structured proteins and hormones. Since intracellular proteins appear to be essentially free of disulphide bonds, a polarization with respect to the 'rough' and 'smooth' sides of the endoplasmic reticulum makes good sense.

The mechanism of action of the interchange enzyme is still not known. Disulphide bonds in proteins appear to occupy the relatively more hydrophobic regions of the molecules. It is possible that a driving force for the series of interchange reactions that lie between the incorrectly and the correctly cross-linked protein is derived from the transfer of solvent-exposed disulphide bridges in the 'scrambled' species to interior positions. Such a purely thermodynamic explanation, based on the transfer of a hydrophobic grouping from an aqueous to a low-dielectric environment, would also be consistent with the non-specificity of the enzyme. The interchange enzyme rearranges the disulphide bonds of molecules as divergent as soya-bean trypsin inhibitor and pancreatic ribonuclease, and its efficiency in the process seems to be correlated with the number of possible combinations of half-cystine residues in the substrate. Soya-bean trypsin inhibitor, with three possible isomeric arrangements of two disulphide bonds, is converted into the native form many times more rapidly than is 'scrambled' ribonuclease, with 105 possible isomers (Steiner *et al.*, 1965).

### *Integrity of protein structure in solution as a highly co-operative process*

Although most proteins, when examined by the standard hydrodynamic or spectroscopic methods, appear to be stable globular objects, so-called 'historical' methods, such as hydrogen exchange or susceptibility to proteolysis, that measure unusual flexibility in solution show clearly that the fabric of the protein molecule is subject to sizable deviations (Schechter *et al.*, 1969; Taniuchi *et al.*, 1969) from the average structure as deduced by crystallography. Such deviations are very infrequent, however, and proteins in solution show a remarkably high degree of co-operative stabilization of conformation. When denaturation does occur, it usually involves a sudden transition, over a narrow range of change in the denaturing agent. Fig. 2 illustrates some selected experiments on nuclease that illustrate this phenomenon: the parameters that were followed as a function of pH during the reversible $H^+$-catalysed transition were intrinsic viscosity and circular dichroism (A. N. Schechter, H. F. Epstein & C. B. Anfinsen, unpublished work; Anfinsen, 1971).

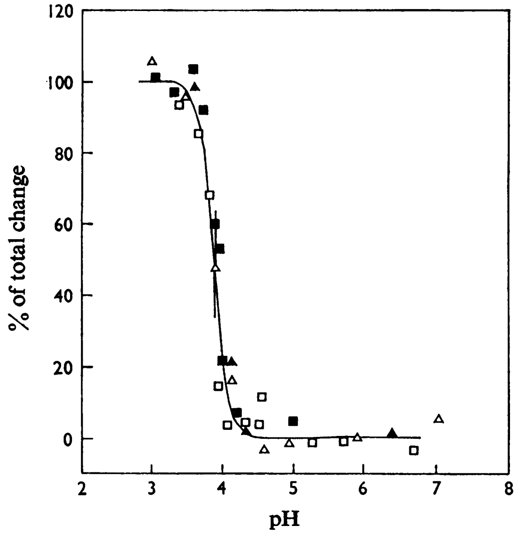


Fig. 2. *Changes in reduced viscosity and molar ellipticity at 220nm during the acid-induced transition from native to denatured nuclease*

□ and ■, Reduced viscosity; △ and ▲, molar ellipticity at 220nm. □ and △, Measurements made during the addition of acid; ■ and ▲, measurements made during the addition of base.

A dramatic experimental example of this co-operativity is found in the case of a derivative of nuclease, containing residues (1–126), that lacks only the 23 *C*-terminal amino acids. This material has optical-rotatory and viscometric properties that are consistent with an essentially random-coil structure, and the inability to form the terminal helix that exists in the native protein approximately between residues 122 and 134 precludes the essential co-operativity required for stability (Taniuchi & Anfinsen, 1969). [Parenthetically, the (1–126) fragment must be able to form a structure containing at least a fleeting resemblance to the active centre of native nuclease, since it possesses about 0.12% the activity of the natural enzyme (D. H. Sachs, H. Taniuchi, A. N. Schechter & A. Eastlake, unpublished work).]

Two kinds of cleavage of the nuclease chain may be made without complete loss of activity and with retention of structure. For example, when nuclease is stabilized to random trypsin digestion by the addition of the ligands thymidine 3′,5′-diphosphate (pdTp) and $Ca^{2+}$, a derivative termed nuclease-T is formed by the restriction of proteolysis to the bonds between residues 5 and 6 and between either residues 48 and 49 or residues 49 and 50 (Fig. 3). This two-fragment derivative exhibits about 8–10% of the activity of native nuclease. The two component fragments, (6–48) and (49–149), or (50–149), are essentially structureless when alone in solution, but a nearly native structure is regenerated in full (except for the missing five or six residues, of course) on reconstitution of the complex (Taniuchi *et al.*, 1967; Taniuchi & Anfinsen, 1968). Plate 1 shows crystals of the complexes of fragment (6–48) with fragment

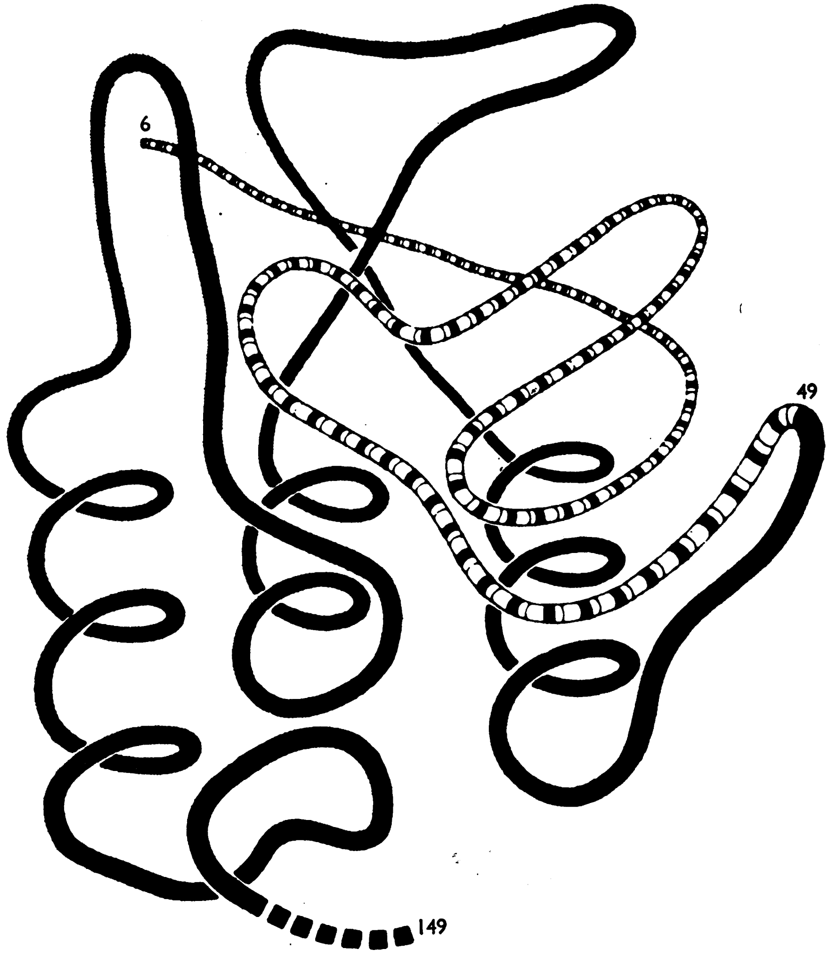
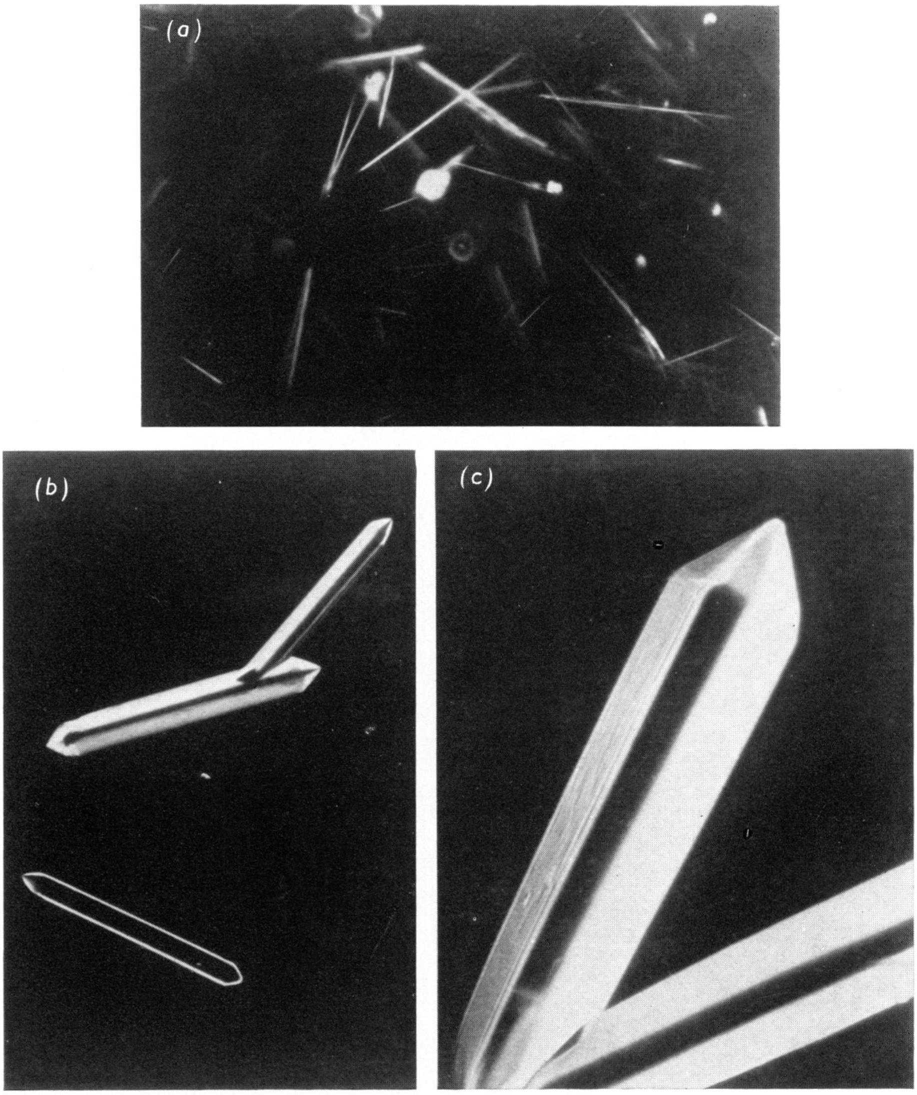


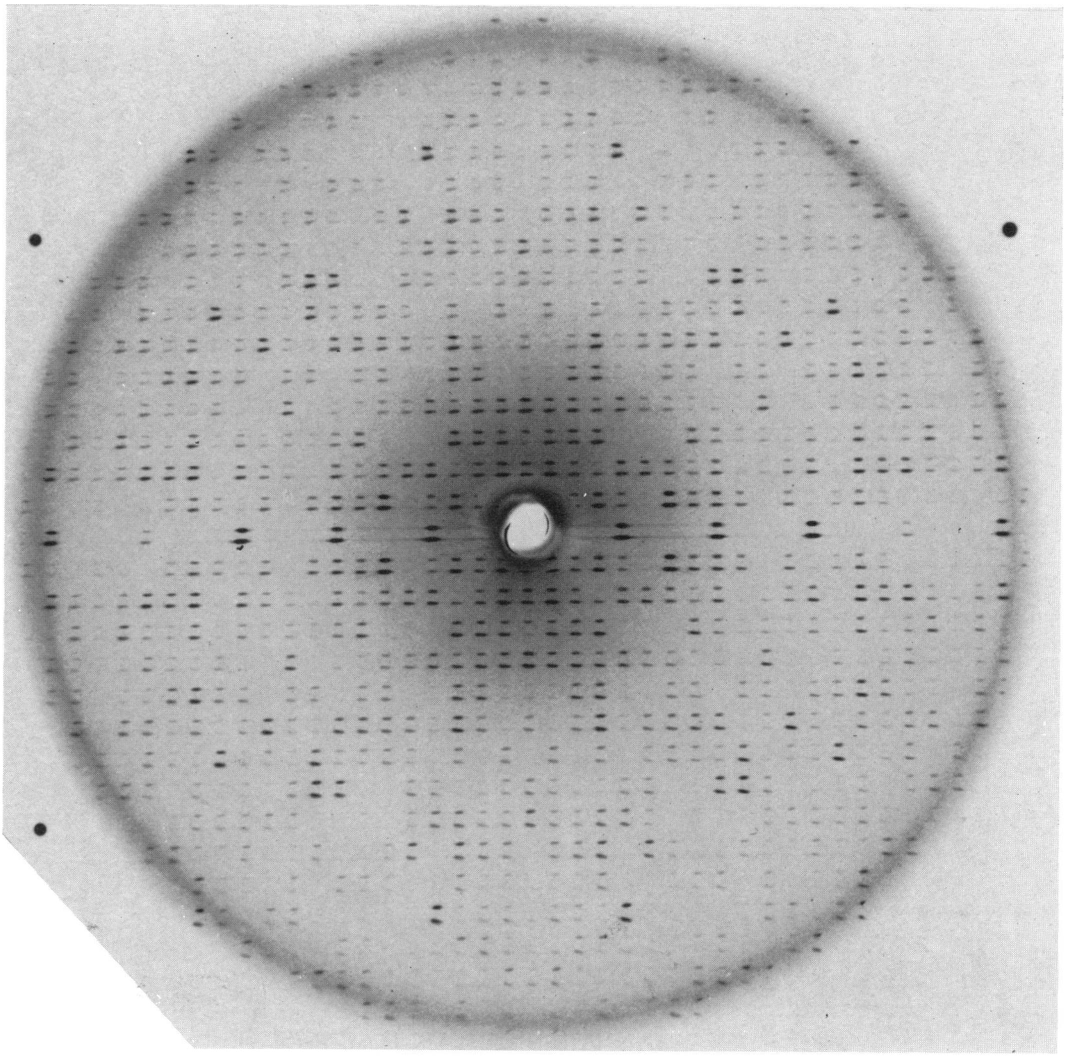
Fig. 3. *Schematic drawing of nuclease-T*



## EXPLANATION OF PLATE I

*Crystalline reconstituted (fragment* (6–48)+*fragment* (50–149)), (*fragment* (6–48)+*fragment* (49–149)) *and reconstituted* (*fragment* (6–48)+*fragment* (49–149)) *of nuclease*

(*a*) Reconstituted (fragment (6–48)+fragment (50–149)); (*b*) (fragment (6–48)+fragment (49–149)); (*c*) reconstituted (fragment (6–48)+fragment(49–149)). The magnification factors of all photographs are the same. For example, the original size of the crystal of reconstituted (fragment (6–48)+(fragment 49–149)) (*c*) was 0.42mm × 0.06mm × 0.06mm. The crystals of (fragment (6–48)+fragment (49–149)) (*b*) and of reconstituted (fragment (6–48)+fragment (49–149)) (*c*) have the prismatic faces of {110} and triangular terminal faces of {111}. Nuclease crystals usually possess prismatic faces of {100} and are terminated by the pyramidal forms {111} (having diamond shapes) and {011} (with triangular faces) (Arnone *et al.*, 1971). However, some crystals of reconstituted (fragment (6–48)+fragment (49–149)) show prismatic faces of both {100} and {110}, although the latter invariably predominate.



EXPLANATION OF PLATE 2

*Comparison of the X-ray-diffraction pattern of a single crystal of liganded nuclease-T′,(fragment* (6–48)+*fragment* (49–149)), *with that of liganded nuclease*

The two (*0kl*) precession photographs (precession angle 15°) of the reflexions are superposed with a slight displacement. The upper rows of spots correspond to the pattern of nuclease-T′.

(49–149) and with fragment (50–149), grown in the presence of thymidine 3′,5-diphosphate and $Ca^{2+}$. The diffraction pattern (Plate 2) given by the liganded (fragment (6–48)+fragment (49–149)) complex (Taniuchi *et al.*, 1972) is extremely similar to that given by liganded native nuclease (Arnone *et al.*, 1971), and the two crystals are isomorphous and fall in the same crystallographic space group. The availability of this crystallizable complex has opened the way for the preparation of semi-synthetic analogues of the enzyme that should be of considerable value in the elucidation of its mechanism of action. Synthesis of analogues of the (6–48) portion is described briefly below.

The absence of structure in the separated fragments, and the formation of native conformation in the reconstituted material, is a striking example of the principle of co-operativity in protein structure. A similar situation exists with a second kind of complementing system derived from nuclease that is composed of a fragment containing residues (99–149) and a second containing residues (1–126) (Taniuchi & Anfinsen, 1971). Once again, synthetic analogues of the *C*-terminal member of the pair have been synthesized and a study of their ability to form three-dimensional structures similar to the native enzyme has been undertaken (Parikh *et al.*, 1971). An approach to the problem of nucleation and folding, by using these fragments, is discussed below in connexion with some recently developed immunochemical probes of structure.

### *Organic synthesis of enzyme analogues*

The use of classical techniques has permitted the total synthesis of a number of large oligopeptides, including adrenocorticotrophin with 39 residues (Schwyzer & Sieber, 1963), followed by thyrocalcitonin (Guttmann *et al.*, 1968) with 32 residues and the two chains of insulin (Katsoyannis *et al.*, 1963; Meienhofer *et al.*, 1963; Wang *et al.*, 1964) with 21 and 30 residues respectively. These molecules were prepared by methods involving the combination of 'blocks' of amino acids that were carefully purified after each amino acid addition. The final products satisfied the criteria that peptide chemists have set for themselves, including such properties as activity, elemental and amino acid analysis, chromatographic purity and a variety of optical and other physical characteristics. However, such methods are relatively indiscriminating with very large molecules, and it becomes difficult to distinguish pure material from a mixture of many very similar structures that might have arisen during terminal deprotection steps or have been introduced as undetected side products during synthesis.

A detailed understanding of the biological function of macromolecules will only truly be achieved when the mechanisms of action of these molecules can be examined at the level of atomic resolution. Since the cellular components with which polypeptide hormones interact are still not known, such direct study of functional geometry is feasible only in the case of enzymes, many of whose three-dimensional structures are now known to a fairly high resolution. Unfortunately, an approach to the problem through the synthesis of analogues by the classical methods that have been employed for the larger hormones is not easily applicable to enzymes, because the use of these methods is restricted at the present time to chains of about 40 or less residues in length owing to problems involving solubility and the sluggishness of coupling of large fragments. The development of the solid-phase method of Merrifield (1965) has introduced a new stimulus to this approach to the study of structure–function relationships. As I shall discuss below, the method, in spite of the inherent difficulties arising from slight incompleteness of coupling at each step, yields materials that can be prepared in a state of purity approaching that of the natural polypeptides when they are susceptible to 'functional' purification (Ontjes & Anfinsen, 1969; Hofmann *et al.*, 1966).

In the case of protein molecules with crystallographically defined three-dimensional structures and active centres, only pancreatic ribonuclease and staphylococcal nuclease permit the preparation of semi-synthetic analogues that can be looked at in terms of the geometry of the analogues relative to the native protein. This is true because these two enzymes have been broken, by subtilisin (Richards, 1958) and trypsin digestion (Taniuchi *et al.*, 1967; Taniuchi & Anfinsen, 1968), into fragments that can recombine through non-covalent interactions with the regeneration of activity and characteristic physical characteristics. The three-dimensional structure of ribonuclease-S (Wyckoff *et al.*, 1970) is essentially the same as that of native ribonuclease (Kartha *et al.* 1967), and conclusions based on investigation of the enzymic properties of semi-synthetic ribonuclease-S should apply directly to the native protein as well. No results have as yet appeared, however, on the crystallographic structures of analogues containing synthetic ribonuclease-S peptide (Hofmann *et al.*, 1966; Scoffone *et al.*, 1967). As mentioned above, Taniuchi and his colleagues have recently shown that nuclease-T is isomorphous with crystals of native nuclease, probably differing only very slightly as a result of the 'cut ends' where the cleavage of the parent chain occurred. In the case of nuclease, therefore, the examination of semi-synthetic analogues having one synthetic and one natural fragment may be possible entirely with the use of difference Fourier analysis, employing the intensities of reflexions from the native and semi-synthetic materials and phase information from the native structure alone. This kind of study will hopefully give us a direct view of 'redesigned' active centres as well as a check on the dependability and accuracy of the solid-phase synthetic method of Merrifield (1965) in the studies of structure and function.

Let me say a few words about means for the preparation of large peptides by the solid-phase method in a state of purity sufficient to permit sound biological conclusions. As we have considered above, several proteins have now been shown to undergo a sort of 'rigidification' when ligands are added that have high affinities for specific sites. Some proteins, like ribonuclease, are already rather immotile, owing, presumably, to the disulphide bridges. The dampening of normal motility of proteins in solution is sufficient to decrease greatly, or in some cases to abolish, general susceptibility to proteolysis (Taniuchi *et al.*, 1969). Synthetic preparations of bovine pancreatic ribonuclease were rid of a great deal of contaminating 'mistake-sequence' impurity by trypsin digestion, and Gutte & Merrifield (1971) were able to isolate a surviving resistant synthetic enzyme that appeared to be about 80% pure by various criteria. Dr. Irwin Chaiken (1971) has capitalized on the resistance of nuclease-T to trypsin in the presence of the ligands thymidine 3′,5′-diphosphate and Ca$^{2+}$ to prepare semi-synthetic nuclease-T, containing synthetic residues (6–47) and natural residues (49–149), which exhibited a purity of well over 90% as judged by chemical, physical and catalytic tests. Similar purifications have been carried out by Parikh *et al.* (1971) on the semi-synthetic complex of the nuclease fragments synthetic fragment [Phe$^{140}$]-(108–145) and natural

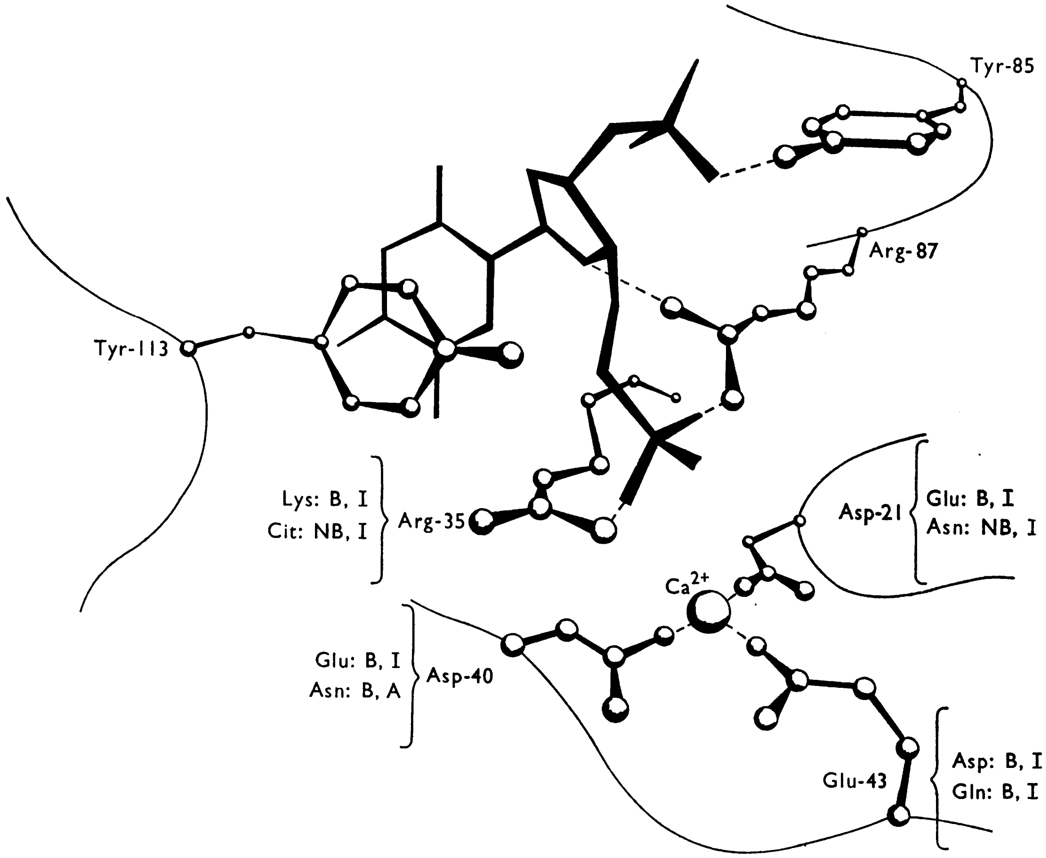


Fig. 4. *Some semi-synthetic nuclease-T derivatives*

The symbols B and NB indicate whether or not the synthetic fragment (6–47) analogue was bound to nuclease fragment (49–149). The symbols I and A indicate whether the resulting complexes were enzymically inactive or active (Chaiken & Anfinsen, 1971).

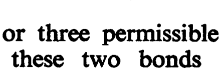
fragment (1–126), and the resulting trypsin-resistant product, after purification on phosphocellulose columns, exhibited a considerable fraction of the activity of the same complex formed from totally natural materials.

The solid-phase method does, at present, involve the inherent difficulty of slight incompleteness of coupling at each step of monomer addition. However, a vigorous international effort is underway to decrease this and other minor deficiencies, and the resulting improvements, together with the use of the 'functional-purification' concept whenever applicable should certainly make the total synthesis of enzymes and their analogues commonplace in the next few years.

I will not relate, in the Lecture, our past efforts on the synthesis of complementing nuclease fragments. These studies, which have involved the preparation of a large variety of active-centre analogues (e.g. see the summary in Fig. 4) (Chaiken & Anfinsen, 1971), derivatives with sequences that have been truncated at either the *C*-terminal or the *N*-terminal end of the polypeptide chain, and sequences with replacement of individual residues that were candidates for involvement in function such as the four histidine residues of the enzyme, have been reported and reviewed elsewhere (see Anfinsen *et al.*, 1971). The analogues are of general interest in relation to the requirements for formation of a functional centre. However, the finding by Taniuchi *et al.* (1972) discussed above, that crystals of liganded nuclease-T and native nuclease are isomorphous and, consequently, that the two forms of the protein are likely to be extremely similar in solution, makes the availability of semi-synthetic analogues of especial value, since the complexes may potentially be examined at atomic resolution, and small changes in bond angles and lengths might be directly related to chemically defined variations in the active centre. Furthermore, as discussed in the section below on immunological study of protein conformation, synthetic variants will also help us to accumulate information on the parameters of structure that govern the spontaneous folding of polypeptide chains and the stabilities of the final products.

### *Nucleation of polypeptide chain folding*

If we permit each of the angles $\psi$ and $\phi$ in the peptide unit:



$$[-C_{\alpha} \overset{\psi}{\curvearrowleft} \overset{\overset{O}{\|}}{C} - \underset{\underset{H}{|}}{N} \overset{\phi}{\curvearrowleft} C_{\alpha} -]$$

to assume two or three permissible values as the atoms forming these two bonds rotate around through 360°, we can calculate that there will exist the possibility of $(2\times2)^{148}$ or $(3\times3)^{148}$ different arrangements of a polypeptide chain of 149 residues, such as staphylococcal nuclease. The time required to test out all of these stereochemical arrangements would clearly be of astronomical proportions, even when we sharply limit the number of permissible angles (as is known to be true from so-called Ramachandran plots of angles in proteins of known structure) and require only a fraction of a nanosecond to try out each possibility.

We know, from kinetic studies of folding, that the approximate half-time for folding of staphylococcal nuclease, for example, is of the order of 0.1–0.2s. There must exist, therefore, certain paths of nucleation and folding that are fairly probable even in the competitive aqueous environment of the protein in solution. The determination of the time required for folding the acid-denatured extended polypeptide chain of nuclease into the native conformation was made by stop-flow measurements of the time required to 'bury' the single tryptophan residue at position 140 in a more hydrophobic environment in the native nuclease structure, as evidenced by a marked enhancement in the intensity of the emission fluorescence of this residue on irradiation of the solution with light at 285nm. A typical fluorescence-enhancement experiment is shown in Fig. 5, where intensity was measured at various pH values, from denaturing values to pH values consistent with intact native structure. The kinetics of the overall process of folding indicated one initial fast step, having a half-time of about 50ms, and a second step (or possibly several overlapping steps) with a longer half-time of about 200ms at room temperature (Fig. 6). Examination of the kinetics as a function of temperature (Fig. 7) showed that the first rapid process was temperature-independent (and therefore

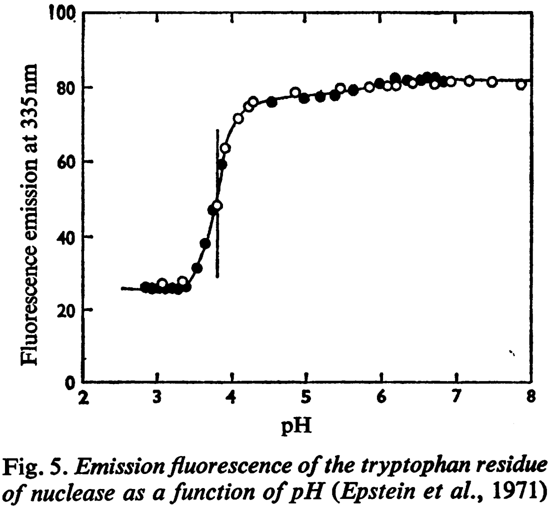


Fig. 5. *Emission fluorescence of the tryptophan residue of nuclease as a function of pH* (*Epstein et al.*, 1971)

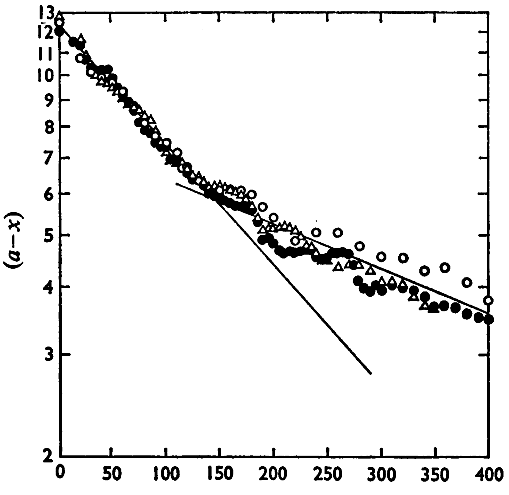


Fig. 6. *First-order kinetic plot of the renaturation of acidified nuclease from pH*3.20 *to pH*6.70

The value of $(a-x)$ is plotted on the ordinate (logarithmic scale) where $a$ is the total increase in fluorescence from the beginning to the observed end of the reaction and $x$ is the increase in fluorescence at any time.

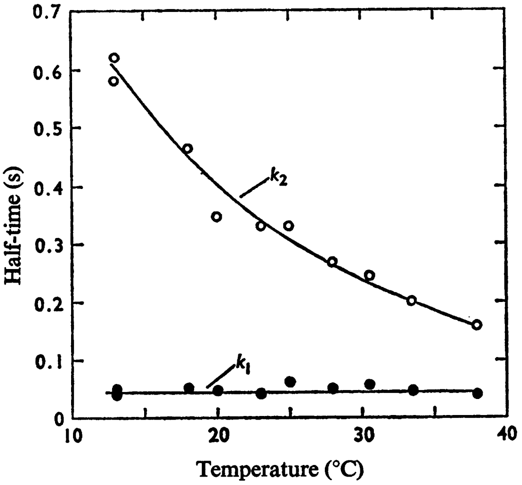


Fig. 7. *Half-time of rates of refolding of acid-denatured nuclease over the temperature range* 13–38°*C*

The symbols $k_1$ and $k_2$ refer to the faster and the slower portions of the refolding process respectively (Epstein *et al.*, 1971).

presumably an entropically driven mechanism), whereas the second showed marked temperature-dependency (Epstein *et al.*, 1971). It is tempting, of course, to attribute the initial entropic step to some coil-to-helix transformation, which would be likely to have such thermodynamic properties. The possible involvement of helix formation as an early step in the nucleation of folding of nuclease is considered below.

A few synthetic poly(amino acids) combine the properties of solubility in water at physiological pH values and organized structure in such a solvent. One example, poly-$N^{\gamma}$-($\omega$-hydroxybutyl)glutamine, which forms a regular right-handed helical structure in neutral aqueous solution, is currently being used by Scheraga and his colleagues in 'host–guest' studies on the effects on helicity of the intermittant introduction of helix-favouring and helix-destroying amino acid residues. The polypeptide chains of protein molecules are, however, singularly free of regular repeating structure. Various portions of a chain, after having assumed a particular position in the native structure, after folding, may be involved in random helical pleated sheet or $\beta$-bend conformations in a helter-skelter way. We have chosen, for the sake of having available a convenient laboratory jargon, to refer to the conformation of a bit of chain in the geometry that it occupies in the native protein as its 'native format'.

The strong co-operativity of structural stabilization shown by proteins, involving a complex matrix of interactions that are of both short-range and long-range character, suggests that most bits of protein chains abstracted from the sequence by cleavage of peptide bonds might have a very limited likelihood of occupying a stable measurable geometry. It is, of course, true that some peptide fragments, such as those derived from the chain of myoglobin that have been studied by Crumpton & Small (1967), Atassi & Singhal (1970), and Epand & Scheraga (1968) and their colleagues, and more recently in this laboratory by Dr. H. Hagenmaier (unpublished work), exhibit a considerable degree of helicity in neutral solution. Our experience with fragments of staphylococcal nuclease (which I suspect might contain a more typical polypeptide chain) indicates that very little stable structure can be formed until almost the entire complement of amino acids is fully assembled.

Fig. 8 summarizes one very general way of looking at the folding problem. We assume here that only limited portions of the chain possess amino acid sequences that are consistent with the formation of co-operatively stabilized structures in solution (labelled 'Random I' and 'Random II' regions). For our purposes here we might think of these as the substance of the two *C*-terminal helices of nuclease that sandwich, between them, the tryptophan residue whose emission-fluorescence properties have been of such diagnostic value in the studies on renaturation from acid solution (see Fig. 5 above).

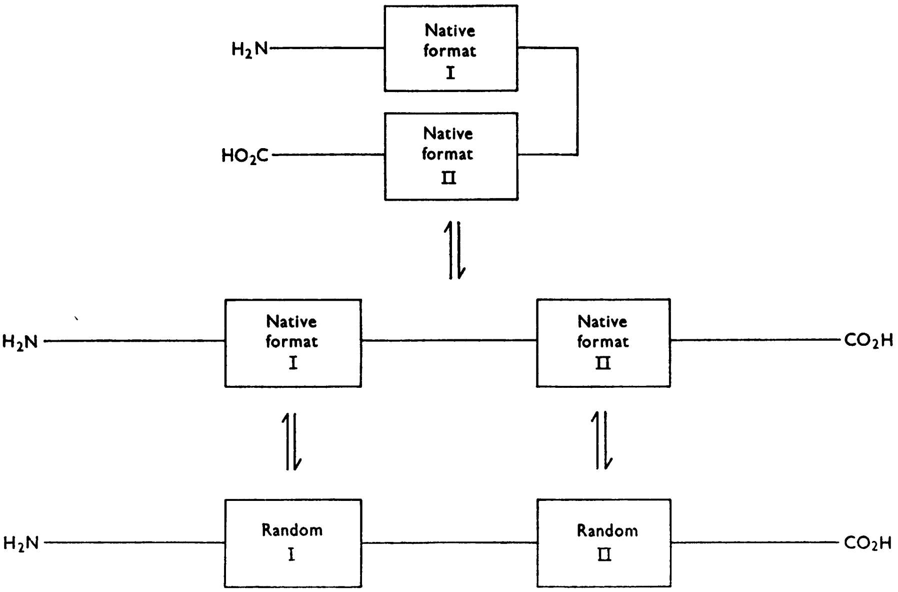

H₂N
Native format I
HO₂C
Native format II
H₂N
Native format I
Native format II
CO₂H
H₂N
Random I
Random II
CO₂H

Fig. 8. *Schematic view of how protein chains might undergo nucleation and folding*

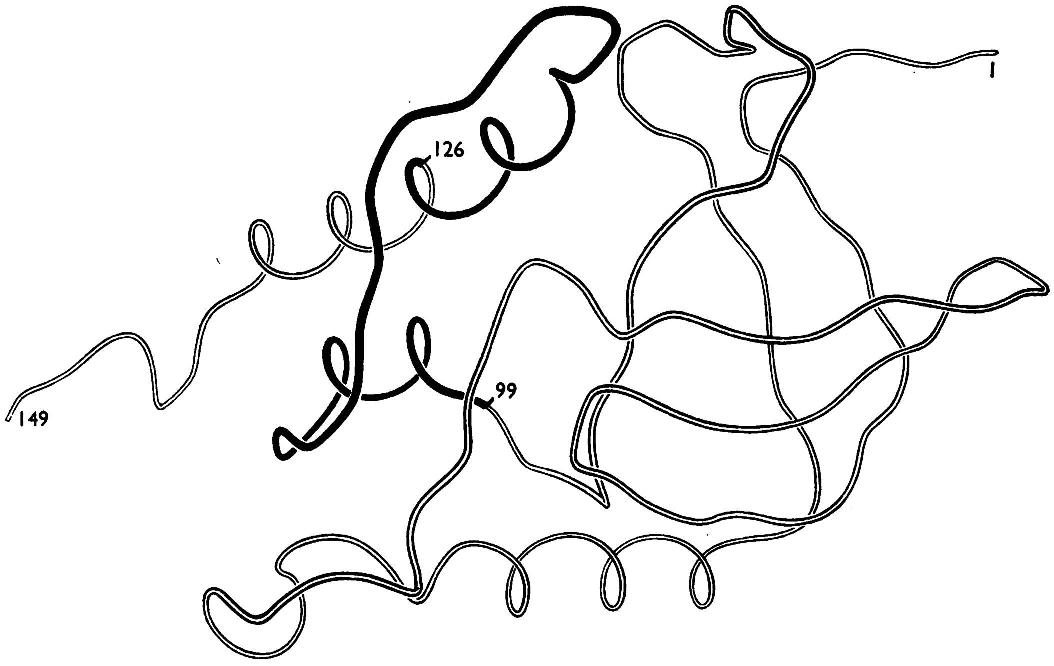


Fig. 9. *View of the backbone of staphylococcal nuclease in three dimensions*

The portion from residue 99 to 149 is cleaved from the rest of the chain by CNBr action on the methionyl residue at position 98. Residues (127–149) are obtained in one fragment during the preparation of the long fragment (1–126) (Taniuchi & Anfinsen, 1969).

Each of these two bits of sequence might be involved in independent flickering equilibria between the random form and the 'native format', i.e. the conformation that this part of the sequence occupies in the native protein structure. When two such fortuitously and transiently organized structures coexist in a single chain, their combination through non-covalent interactions might create a 'nucleation site' and would be the focus for a rapid cascade of interactions leading to the native structure.

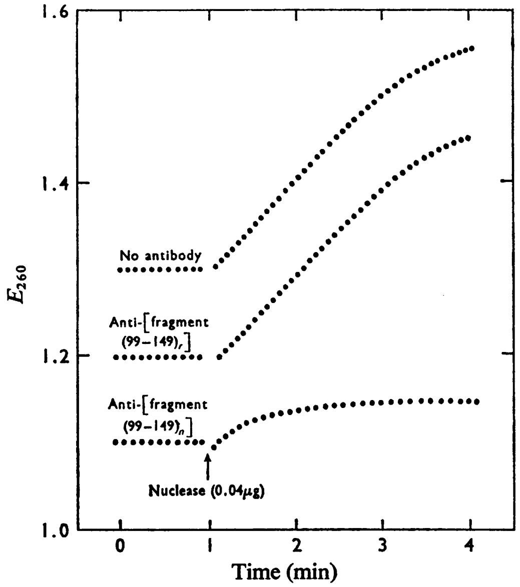


Fig. 10. *Inhibition of nuclease activity by anti-[fragment* (99–149)$_n$] *(antibody fraction isolated by immunoadsorption from a population of antibodies made against the native protein) and the lack of inhibition by the antibody, anti-[fragment* (99–149)$_r$], *made against the peptide* (99–149), *presumably in a random conformation (D. H. Sachs, A. N. Schechter, A. Eastlake & C. B. Anfinsen, unpublished work)*

The direct physical measurement of the folded fraction of the sequence in question, whether still part of the intact protein chain or in the form of fragments such as those complementing oligopeptides from nuclease and ribonuclease that we have considered above, is not feasible at present except in a conveniently helical protein like myoglobin. It is possible that further improvements in the techniques of circular dichroism, nuclear magnetic resonance or laser Raman spectra might eventually permit the use of stop-flow methods for the study of such flickering equilibria. At the moment, however, we must resort to indirect methods. One approach that has been employed recently in our laboratory (D. H. Sachs, A. N. Schechter, A. Eastlake & C. B. Anfinsen, unpublished work) involves the use of antibodies as analytical tools for the detection of the folded and unfolded forms of proteins and polypeptide fragments.

If the total goat anti-nuclease γ-globulin, isolated on a Sepharose–nuclease immunoabsorbent column, is subfractionated on affinity-chromatography columns bearing peptide fragments of nuclease, it is possible to prepare site-specific antibodies that presumably recognize the immobilized fragments only when they 'flicker' into the native formats against which the antibody molecules were originally formed (so-called 'conformational determinants', Sela *et al.*, 1967). The drawing in Fig. 9 shows the (99–149) region of the sequence of nuclease, which contributes the material of the two terminal helices

Table 1. *Results from experiments on the determination of the* $k_{on}$ *rate for the reaction between anti-[fragment* (99–149)$_n$] *and nuclease*

The $k_{off}$ rate is negligible, since it has been shown in experiments, not detailed here, that the equilibrium constant for the complex is of the order of $10^{-8}$. Abbreviation: Ab, antibody.

$$k_{on} = \frac{\ln 2}{t_{\frac{1}{2}} \cdot [\mathrm{Ab}]}$$

| Amount of Ab (μl) | $[\mathrm{Ab}]_{\mathrm{sites}}$ (μM) | $t_{\frac{1}{2}}$(s) | $10^{-5} \times k_{on}$ ($\mathrm{M}^{-1} \cdot \mathrm{s}^{-1}$) |
|---|---|---|---|
| 0 | 0 | ∞ | — |
| 5 | 0.0381 | 46 | 3.95 |
| 5 | 0.0381 | 45 | 4.04 |
| 10 | 0.0762 | 22 | 4.13 |
| 10 | 0.0762 | 23 | 3.96 |
| 20 | 0.152 | 11 | 4.15 |
| 20 | 0.152 | 11 | 4.15 |

$k_{on} = 4.1 \pm 0.1 \times 10^5 \mathrm{M}^{-1} \cdot \mathrm{s}^{-1}$

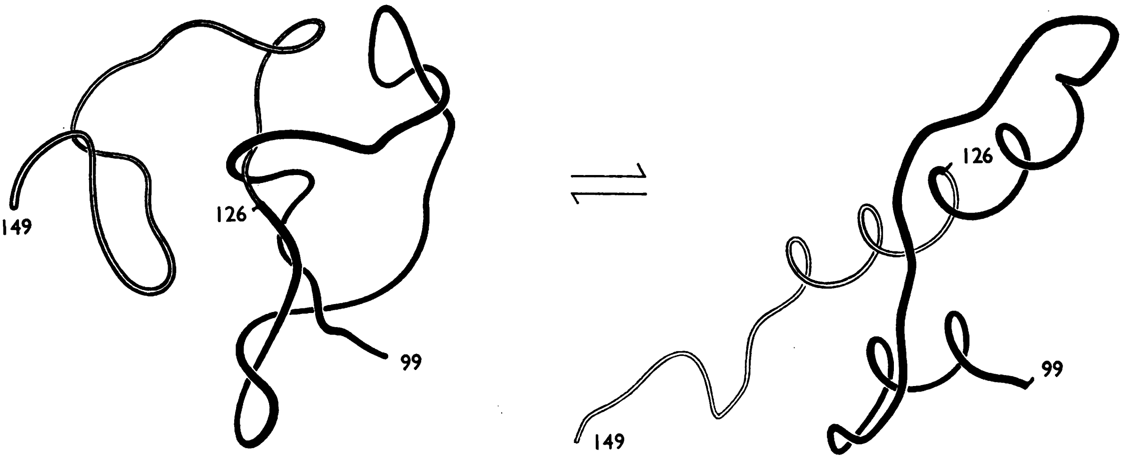


Fig. 11. *Schematic representation of the postulated equilibrium between fragment* (99–149) *of nuclease in its predominantly random form and that fraction of the population in the 'native format'*

of this protein. When the antibody fraction that was isolated by adsorption on Sepharose–fraction (99–149) is added to nuclease, enzyme activity is rapidly neutralized and the rate of loss of activity is a function of the amount of antibody added and is independent of nuclease concentration in antibody excess (Fig. 10). Fig. 10 also shows that antibody made against free fragment (99–149) (abstracted from the protein by CNBr treatment of the native nuclease chain) does not cause inactivation. The inactivating effect, therefore, is a function of the ability of the antibody to recognize the 'native format' of the corresponding sequence of nuclease. Analysis of the kinetics of inactivation yields a value for the $k_{\mathrm{on}}$ of the antibody–nuclease combination as shown in Table 1.

By sedimentation-velocity studies, anti-[fragment $(99\text{–}149)_n$] appears to be multivalent and capable of formation of an antigen–antibody microaggregate. Passage of the anti-[fragment $(99\text{–}149)_n$] antibody through a second Sepharose column bearing attached fragment (127–149) leads to the adsorption of a fraction of the total material. The fraction that passes through the column, anti-[fragment $(99\text{–}126)_n$], is univalent as estimated by the same ultracentrifugal technique. It retains the inactivating capacity of the original globulin fraction and is directed, in its specificity, against a determinant in the fraction of the nuclease fragment shown in Fig. 11 in black. Fig. 11 depicts fragment (99–149) undergoing a reversible transformation from random to 'native-format' geometry.

We may set up the equilibria shown in Scheme 1, the first of which relates random and 'native-format' fragment (99–149) (called E in Scheme 1), and the second of which relates to the interaction between the folded form of fragment (99–149) and anti-[fragment $(99\text{–}126)_n$] antibody. In the latter equilibrium the amount of unbound antibody may be estimated by the rate of loss of activity of a sample of nuclease added to the preincubated mixture of fragment (99–149) and anti-[fragment $(99\text{–}126)_n$]. By combining the two equilibria in one equation, determining the amount of unbound antibody by activity measurements in the presence of increasing amounts of total fragment (99–149), and making the (only) assumption that $K_{\mathrm{assoc.}}$ for the combination of anti-[fragment $(99\text{–}126)_n$] with fragment (99–149) in its native format is the same as that for anti-[fragment $(99\text{–}126)_n$] with nuclease, it is possible to calculate, as summarized in Table 2, that approx. 0.02% of fragment (99–149) exists in the 'native format' at any time.

$$\mathrm{E_r} \rightleftharpoons \mathrm{E_n} \qquad K_{\mathrm{conf.}} = \frac{[\mathrm{E_n}]}{[\mathrm{E_r}]}$$

$$\mathrm{Ab} + \mathrm{E_n} \rightleftharpoons \mathrm{Ab\text{–}E_n} \qquad K_{\mathrm{assoc.}} = \frac{[\mathrm{Ab\text{–}E_n}]}{[\mathrm{Ab}]\cdot[\mathrm{E_n}]}$$

$$K_{\mathrm{conf.}} = \frac{[\mathrm{Ab\text{–}E_n}]}{K_{\mathrm{assoc.}}\cdot[\mathrm{Ab}]\cdot[\mathrm{E_r}]}$$

Scheme 1. *Two equilibria involving fragment* (99–149) *of nuclease with the corresponding equilibrium-constant expressions*

Abbreviations: E, fragment (99–149); Ab, antibody.

Table 2. *Determination of the equilibrium constant for the reaction involving the interconversion of the peptide fragment* (99–149) *in its random form* [*fragment* $(99\text{–}149)_r$] *and in the form this fragment assumes in the native structure of nuclease* [*fragment* $(99\text{–}149)_n$]

Abbreviations: E, fragment (99–149); Ab, antibody.

$$K_{\mathrm{conf.}} = \frac{[\mathrm{Ab\text{–}E_n}]}{K_{\mathrm{assoc.}}\cdot[\mathrm{Ab}]\cdot[\mathrm{E_T}]}$$

| $[\mathrm{Ab}]_{\mathrm{total\ sites}}$ ($\mu$M) | $[\mathrm{E_T}]$ ($\mu$M) | $t_{\frac{1}{2}}$(s) | $[\mathrm{Ab}]_{\mathrm{free\ sites}}$ ($\mu$M) | $[\mathrm{Ab}]_{\mathrm{bound\ sites}}$ ($\mu$M) | $K_{\mathrm{conf.}}$ | % of $\mathrm{E_T}$ as $\mathrm{E_n}$ |
|---|---|---|---|---|---|---|
| 0.076 | 0 | 18 | 0.076 | 0 | — | — |
| 0.076 | 0.65 | 20 | 0.068 | 0.0080 | $2.20\times10^{-4}$ | 0.022 |
| 0.076 | 2.0 | 24 | 0.057 | 0.019 | $2.02\times10^{-4}$ | 0.020 |
| 0.076 | 2.6 | 27 | 0.051 | 0.025 | $2.29\times10^{-4}$ | 0.023 |
| 0.076 | 7.8 | 35 | 0.039 | 0.037 | $1.47\times10^{-4}$ | 0.015 |
| 0.076 | 6.5 | 33 | 0.042 | 0.034 | $1.51\times10^{-4}$ | 0.015 |

$$K_{\mathrm{conf.}} = (2.0\pm0.4)\times10^{-4}$$

I have presented only a bare outline of the immunochemical approach to the study of unfavourable equilibria such as probably exist for many abstracted peptide fragments in solution. This technique, may permit us to obtain some of the quantitative kinetic values that will be needed in the eventual understanding of the nucleation and folding of polypeptide chains.

## References

Anfinsen, C. B. (1971) *Proc. Int. Congr. Pure Appl. Chem. 23rd.* 7, 263

Anfinsen, C. B. & Haber, E. (1961) *J. Biol. Chem.* **236**, 1361

Anfinsen, C. B., Haber, E., Sela, M. & White, F. H., Jr. (1961) *Proc. Nat. Acad. Sci. U.S.* **47**, 1309

Anfinsen, C. B., Cuatrecasas, P. & Taniuchi, H. (1971) *Enzymes*, 3rd. edn., **4**, 177

Anson, M. L. (1945) *Advan. Protein Chem.* **2**, 361

Arnone, A., Bier, C. J., Cotton, F. A., Hazen, E. E., Jr., Richardson, D. C. & Richardson, J. S. (1969) *Proc. Nat. Acad. Sci. U.S.* **64**, 420

Arnone, A., Bier, C. J., Cotton, F. A., Hazen, E. E., Jr., Richardson, D. C., Richardson, J. S. & Yonath, A. (1971) *J. Biol. Chem.* **246**, 2302

Atassi, M. Z. & Singhal, R. P. (1970) *J. Biol. Chem.* **245**, 5122

Bohnert, J. L. & Taniuchi, H. (1972) *J. Biol. Chem.* in the press

Canfield, R. E. & Anfinsen, C. B. (1963) *Biochemistry* **2**, 1073

Chaiken, I. M. (1971) *J. Biol. Chem.* **246**, 2948

Chaiken, I. M. & Anfinsen, C. B. (1971) *J. Biol. Chem.* **246**, 2285

Crumpton, M. J. & Small, P. A. (1967) *J. Mol. Biol.* **26**, 143

DeLorenzo, F., Goldberger, R. F., Steers, E., Givol, D. & Anfinsen, C. B. (1966) *J. Biol. Chem.* **241**, 1562

Dintzis, H. M. (1961) *Proc. Nat. Acad. Sci. U.S.* **47**, 247

Epand, R. M. & Scheraga, H. A. (1968) *Biochemistry* **7**, 2864

Epstein, C. J., Goldberger, R. F. & Anfinsen, C. B. (1963) *Cold Spring Harbor Symp. Quant. Biol.* **28**, 439

Epstein, H. F., Schechter, A. N., Chen, R. F. & Anfinsen, C. B. (1971) *J. Mol. Biol.* **60**, 499

Fuchs, S., DeLorenzo, F. & Anfinsen, C. B. (1967) *J. Biol. Chem.* **242**, 398

Givol, D., DeLorenzo, F., Goldberger, R. F. & Anfinsen, C. B. (1965) *Proc. Nat. Acad. Sci. U.S.* **53**, 766

Goldberger, R. F., Epstein, C. J. & Anfinsen, C. B. (1963) *J. Biol. Chem.* **238**, 628

Gutte, B. & Merrifield, R. B. (1971) *J. Biol. Chem.* **246**, 1922

Guttmann, St., Pless, J., Sandrin, Ed., Jaquenoud, P.-A., Bossert, H. & Willems, H. (1968) *Helv. Chim. Acta* **51**, 1155

Hofmann, K., Finn, F. M., Linetti, M., Montibeller, J. & Zanetti, G. (1966) *J. Amer. Chem. Soc.* **88**, 3633

Kartha, G., Bello, J. & Harker, D. (1967) *Nature* (*London*) **213**, 862

Katsoyannis, P. G., Tometsko, A. & Fukoda, K. (1963) *J. Amer. Chem. Soc.* **85**, 2863

Khorana, H. G. (1968) *Biochem. J.* **109**, 709

Lumry, R. & Eyring, H. (1954) *J. Phys. Chem.* **58**, 110

Meienhofer, J., Schnabel, E., Bremer, H., Brinkhoff, O., Zabel, R., Sroka, W., Klostermeyer, H., Brandenburg, D., Okuda, T. & Zahn, H. (1963) *Z. Naturforsch.* **186**, 1120

Merrifield, R. B. (1965) *Science* **150**, 178

Ontjes, D. & Anfinsen, C. B. (1969) *J. Biol. Chem.* **244**, 6316

Parikh, I., Corley, L. & Anfinsen, C. B. (1971) *J. Biol. Chem.* **246**, 7392

Rabin, B. R., Sunshine, G. H. & Williams, D. J. (1970) *Biochem. Soc. Symp.* **31**, 203

Richards, F. M. (1958) *Proc. Nat. Acad. Sci. U.S.* **44**, 162

Schechter, A. N., Moravek, L. & Anfinsen, C. B. (1968) *Proc. Nat. Acad. Sci. U.S.* **61**, 1478

Schechter, A. N., Moravek, L. & Anfinsen, C. B. (1969) *J. Biol. Chem.* **244**, 4981

Schwyzer, R. & Sieber, P. (1963) *Nature* (*London*) **199**, 172

Scoffone, E., Rocchi, R., Marchiori, F., Moroder, L., Marzotto, A. & Tamburro, A. M. (1967) *J. Amer. Chem. Soc.* **89**, 5450

Sela, M., White, F. H. & Anfinsen, C. B. (1957) *Science* **125**, 691

Sela, M., Schechter, B., Schechter, I. & Borek, F. (1967) *Cold Spring Harbor Symp. Quant. Biol.* **32**, 537

Steiner, R. F., DeLorenzo, F. & Anfinsen, C. B. (1965) *J. Biol. Chem.* **240**, 4648

Taniuchi, H. & Anfinsen, C. B. (1968) *J. Biol. Chem.* **243**, 4778

Taniuchi, H. & Anfinsen, C. B. (1969) *J. Biol. Chem.* **244**, 38

Taniuchi, H. & Anfinsen, C. B. (1971) *J. Biol. Chem.* **246**, 2291

Taniuchi, H., Anfinsen, C. B. & Sodja, A. (1967) *Proc. Nat. Acad. Sci. U.S.* **58**, 1235

Taniuchi, H., Moravek, L. & Anfinsen, C. B. (1969) *J. Biol. Chem.* **244**, 4600

Taniuchi, H., Davies, D. & Anfinsen, C. B. (1972) *J. Biol. Chem.* in the press

Venetianer, P. & Straub, F. B. (1963) *Biochim. Biophys. Acta* **67**, 166

Venetianer, P. & Straub, F. B. (1964) *Biochim. Biophys. Acta* **89**, 189

Wang, Y., Hsu, J.-Z., Chang, W.-C., Cheng, L.-L., Hsing, C.-Y., Chi, A.-H., Loh, T.-P., Li, C.-H., Shi, P.-T. & Yieh, Y.-H. (1964) *Sci. Sinica* **13**, 2030

White, F. H., Jr. (1961) *J. Biol. Chem.* **236**, 1353

Wyckoff, H. W., Tsernoglou, D., Hanson, A. W., Knox, J. R., Lee, B. & Richards, F. M. (1970) *J. Biol. Chem.* **245**, 305

Zamecnik, P. (1962) *Biochem. J.* **85**, 257